Attorney Docket No.15473US01

SYSTEM AND METHOD FOR TIME DOMAIN AUDIO SLOW DOWN, WHILE MAINTAINING PITCH

RELATED APPLICATIONS

[0001]   This application makes reference to Manoj Kumar Singhal, *et al.* United States Non-Provisional Application Serial No. _____ (Attorney Docket No. 15474US01) entitled "System and Method for Time Domain Audio Speed Up, While Maintaining Pitch" filed March 18, 2004, the complete subject matter of which is hereby incorporated herein by reference, in its entirety.

[0002]   Reference is also made to Manoj Kumar Singhal, *et al.* United States Non-Provisional Application Serial No. _____ (Attorney Docket No. 15475US01) entitled "System and Method for Frequency Domain Audio Speed Up or Slow Down, While Maintaining Pitch" filed March 18, 2004, the complete subject matter of which is hereby incorporated herein by reference, in its entirety.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0003]   [Not Applicable]

[MICROFICHE/COPYRIGHT REFERENCE]

[0004]   [Not Applicable]

BACKGROUND OF THE INVENTION

[0005]   In many audio applications, an audio signal may be modified or processed to achieve a desired characteristic or quality. One of the characteristics of

1

an audio signal that is frequently processed or modified is the speed of the signal at which it needs to be played. When sounds are recorded, they are often recorded at the normal speed and frequency at which the source plays or produces the signal. When the speed of the signal is modified, however, the frequency often changes, which may be noticed in a changed pitch and the voice does not resemble with the original signal. For example, if the voice of a woman is recorded at a normal level but played back at a slightly slower rate, the woman's voice will resemble that of a man, or a voice at a lower frequency. Similarly, if the voice of a man is recorded at a normal level then played back at a slightly faster rate, the man's voice will resemble that of a woman, or a voice at a higher frequency.

[0006] Some applications may require that an audio signal be played at a slower rate, while maintaining the same original frequency, i.e. keeping the pitch of the sound at the same value as when played back at the normal speed.

[0007] Further limitations and disadvantages of conventional and traditional approaches will become apparent to one of ordinary skill in the art through comparison of such systems with the present invention as set forth in the remainder of the present application with reference to the drawings.

BRIEF SUMMARY OF THE INVENTION

[0008] Aspects of the present invention may be seen in a method for slowing down an encoded original audio signal, said original audio signal having an original frequency and original playback speed. The method being done in a system with a machine-readable storage having stored thereon, a computer program having at least one code section. The at least one code section being executable by a machine for causing the machine to perform operations comprising receiving the encoded original audio signal; retrieving frames of the original audio signal; replicating frames at a rate according to a desired playback speed; wherein said desired playback speed is less than the original playback speed; applying a window function to the replicated frames; converting the signal with replicated frames from digital to analog format; and using the original frequency to playback the analog format signal.

[0009] The system comprises at least one processor capable of receiving the encoded original audio signal; retrieving frames of the original audio signal; replicating frames at a rate according to a desired playback speed; applying a window function to the replicated frames; converting the signal with replicated frames from digital to analog format; and using the original frequency to playback the analog format signal.

[0010] The method comprises receiving the encoded original audio signal; retrieving frames of the original audio signal; replicating frames at a rate according to a desired playback speed; applying a window function to the replicated frames; converting the signal with replicated

3

frames from digital to analog format; and using the original frequency to playback the analog format signal.

[0011]   In an embodiment of the present invention, the desired playback speed is a predefined default value.

[0012]   In another embodiment of the present invention, the desired playback speed is a programmable value.

[0013]   These and other features and advantages of the present invention may be appreciated from a review of the following detailed description of the present invention, along with the accompanying figures in which like reference numerals refer to like parts throughout.

BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

[0014]   **Figure 1** illustrates a block diagram of an exemplary time-domain encoding of an audio signal, in accordance with an embodiment of the present invention.

[0015]   **Figure 2** illustrates a block diagram of an exemplary time-domain decoding of an audio signal, in accordance with an embodiment of the present invention.

[0016]   **Figure 3** illustrates a flow diagram of an exemplary method for time-domain decoding of an audio signal, in accordance with an embodiment of the present invention.

[0017]   **Figure 4** illustrates a block diagram of an exemplary frequency-domain encoding of an audio signal, in accordance with an embodiment of the present invention.

[0018]   **Figure 5** illustrates a block diagram of an exemplary frequency-domain decoding of an audio signal, in accordance with an embodiment of the present invention.

[0019]   **Figure 6** illustrates a flow diagram of an exemplary method for frequency-domain decoding of an audio signal, in accordance with an embodiment of the present invention.

[0020]   **Figure 7** illustrates a block diagram of an exemplary audio decoder, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0021]   The present invention relates generally to audio decoding.   More specifically, this invention relates to decoding audio signals to obtain an audio signal at a slower speed while maintaining the same pitch as the original audio signal.   Although aspects of the present invention are presented in terms of a generic audio signal, it should be understood that the present invention may be applied to many other types of systems.

[0022]   **FIGURE 1** illustrates a block diagram of an exemplary time-domain encoding of an audio signal 111, in accordance with an embodiment of the present invention. The audio signal 111 is captured and sampled to convert it from analog-to-digital format using, for example, an audio to digital converter (ADC).   The samples of the audio signal 111 are then grouped into frames 113 ($F_0 \ldots F_n$) of 1024 samples such as, for example, ($F_x(0) \ldots F_x(1023)$).   The frames 113 are then encoded according to one of many encoding schemes depending on the system.

[0023]   **Figure 2** illustrates a block diagram of an exemplary time-domain decoding of an audio signal, in accordance with an embodiment of the present invention.   In an embodiment of the present invention, the input to the decoder is frames 213 ($F_0 \ldots F_n$) of 1024 samples such as, for example, frames 113 ($F_0 \ldots F_n$) of 1024 samples of **Fig. 1**.

[0024]   The frames 213 ($F_0 \ldots F_n$) are then replicated at a rate consistent with the desired slow rate. For example, if the desired audio speed is half the original speed, then each frame is repeated, resulting in frames 212 ($FR_0 \ldots FR_m$) of 1024 samples, where $FR_0 = FR_1 = F_0$, and $FR_2 = FR_3 = F_1$, etc. Additionally, m depends on the desired slow rate. In

6

the example, where the desired audio speed is half the original speed, m = 2n. If, for example, the desired audio speed is two-thirds of the original speed, then every alternate frame is repeated, so frames 213 $(F_0...F_n)$ result in frames $(FR_0...FR_m)$, where $FR_0 = F_0$, $FR_1 = FR_2 = F_1$, $FR_3 = F_2$, $FR_4 = FR_5 = F_3$, etc., and m = 3n/2. So, the same argument can be extended to support any speed between the input and output signal once the speed ratio is computed. So, the idea is to generate "u" frames from "v" frames for a given "v/u" speed ratio.

[0025]   A window function WF is then applied to frames 212 $(FR_0...FR_m)$ to "smooth out" the samples and ensure that the resulting signal does not have any artifacts that may result from repeating each frame. The window function results in the windowed frames 214 $(WF_0... WF_L)$ of 1024 samples. The window function WF can be one of many widely known and used window functions, or can be designed to accommodate the design requirements of the system.

[0026]   The windowed frames 214 $(WF_0... WF_L)$ of 1024 samples are then run through a digital-to-analog converter (DAC) to get an analog signal 201. The analog signal 211 is a longer version of the analog input signal 111 of **Fig. 1** (analog signal 211 and analog signal 111 are not equal). When the analog signal 211 is played at the same frequency as the original signal 111 of **Fig. 1**, the speed, in the example with repeating each frame, is effectively half the speed at which the original audio was but the pitch remains the same, since the playback frequency remains unchanged. Hence, achieving a slower audio playback without affecting the pitch.

[0027]   **Figure** 3 illustrates a flow diagram of an exemplary method for time-domain decoding of an audio

7

signal, in accordance with an embodiment of the present invention. At a starting block 421, an input is received from the encoder directly, using a storage device, or through a communication medium. The input, which is coming from the encoder, is frames ($F_0...F_n$). Then depending on the rate at which the audio signal needs to be slowed down, the proper number of frames is replicated at a next block 423, as described above with reference to **Fig. 2**, resulting in the replicated frames ($FR_0...FR_m$).

[0028] At a next block 425, a window function WF is applied to the frames ($FR_0...FR_m$) to "smooth out" the samples and ensure that the resulting signal does not have any artifacts that may result from repeating each frame. The window function results in the windowed frames ($WF_0...WF_L$). The window function WF can be one of many widely known and used window functions like Hanning, Hamming, Blackman or Gaussian. The choice of the window function depending upon the property of the windows or a specific window can be designed to accommodate the design requirements of the system.

[0029] The windowed frames ($WF_0... WF_L$) are then sent through the DAC at a next block 427 to produce the audio signal at the desired slower speed, with the same pitch as the original because the playback frequency is kept the same as the original signal.

[0030] Standards such as, for example, MPEG-1, Layer 3 (MPEG stands for Motion Pictures Experts Group), MPEG-4 Advance Audio Coding (AAC) or Dolby Digital AC-3 decoder have been devised for compressing audio signals. In certain embodiments of the present invention, the audio signal can be compressed in accordance with such standards for compressing audio signals.

8

[0031] **Figure 4** illustrates a block diagram describing the encoding of an audio signal 101, in accordance with the MPEG-1, Layer 3 standard, MPEG-4 AAC or Dolby Digital AC-3 decoder. The audio signal 101 is captured and used for further audio post processing depending upon the speed. The samples of the audio signal 101 are then grouped into frames 103 $(F_0 \ldots F_n)$ of 1024 samples such as, for example, $(F_x(0) \ldots F_x(1023))$.

[0032] The frames 103 $(F_0 \ldots F_n)$ are then grouped into windows 105 $(W_0 \ldots W_n)$ each one of which comprises 2048 samples or two frames such as, for example, $(W_x(0) \ldots W_x(2047))$ comprising frames $(F_x(0) \ldots F_x(1023))$ and $(F_{x+1}(0) \ldots F_{x+1}(1023))$. However, each window 105 W_x has a 50% overlap with the previous window 105 W_{x-1}. Accordingly, the first 1024 samples of a window 105 W_x are the same as the last 1024 samples of the previous window 105 W_{x-1}. For example, $W_0 = (W_0(0) \ldots W_0(2047)) = (F_0(0) \ldots F_0(1023))$ and $(F_1(0) \ldots F_1(1023))$, and $W_1 = (W_1(0) \ldots W_1(2047)) = (F_1(0) \ldots F_1(1023))$ and $(F_2(0) \ldots F_2(1023))$. Hence, in the example, W_0 and W_1 contain frames $(F_1(0) \ldots F_1(1023))$.

[0033] A window function w(t) is then applied to each window 105 $(W_0 \ldots W_n)$, resulting in sets $(wW_0 \ldots wW_n)$ of 2048 windowed samples 107 such as, for example, $(wW_x(0) \ldots wW_x(2047))$. A Modified Discrete Cosine or Fourier Transform (MDCT/FT) is then applied to each set $(wW_0 \ldots wW_n)$ of windowed samples 107 $(wW_x(0) \ldots wW_x(2047))$, resulting sets $(MDCT_0 \ldots MDCT_n)$ of 1024 frequency coefficients 109 such as, for example, $(MDCT_x(0) \ldots MDCT_x(1023))$.

[0034] The sets of frequency coefficients 109 $(MDCT_0 \ldots MDCT_n)$ are then quantized and coded for transmission, forming an audio elementary stream (AES). The AES can be multiplexed with other AESs. The multiplexed

signal, known as the Audio Transport Stream (Audio TS) can then be stored and/or transported for playback on a playback device. The playback device can either be at a local or remote located from the encoder. Where the playback device is remotely located, the multiplexed signal is transported over a communication medium such as, for example, the Internet. The multiplexed signal can also be transported to a remote playback device using a storage medium such as, for example, a compact disk.

[0035] During playback, the Audio TS is de-multiplexed, resulting in the constituent AES signals. The constituent AES signals are then decoded, yielding the audio signal. During playback the speed of the signal may be decreased to produce the original audio at a slower speed.

[0036] **Figure 5** is a block diagram describing the decoding of an audio signal, in accordance with another embodiment of the present invention. In an embodiment of the present invention, the input to the decoder is sets $(MDCT_0...MDCT_n)$ of 1024 frequency coefficients 209 such as, for example, the sets $(MDCT_0...MDCT_n)$ of 1024 frequency coefficients 109 of **Fig. 4**. An inverse modified discrete cosine transform (IMDCT) is applied to each set $(MDCT_0...MDCT_n)$ of 1024 frequency coefficients 209. The result of applying the IMDCT is the sets $(wW_0...wW_n)$ of windowed samples 207 $(wW_x(0)...wW_x(2047))$ equivalent to sets $(wW_0...wW_n)$ of windowed samples 107 $(wW_x(0)...wW_x(2047))$ of **Fig. 4**.

[0037] An inverse window function $w_I(t)$ is then applied to each set $(wW_0...wW_n)$ of 2048 windowed samples 207, resulting in windows 205 $(W_0...W_n)$ each one of which comprises 2048 samples. Each window 205 $(W_0...W_n)$ comprises 2048 samples from two frames such as, for example,

$(W_x(0)...W_x(2047))$ comprising frames $(F_x(0)...F_x(1023))$ and $(F_{x+1}(0)...F_{x+1}(1023))$ as illustrated in **Fig. 4**. The frames 203 $(F_0...F_n)$ of 1024 samples such as, for example, $(F_x(0)...F_x(1023))$, are then extracted from the windows 205 $(W_0...W_n)$.

[0038] The frames 203 $(F_0...F_n)$ are then replicated at a rate consistent with the desired slow rate. For example, if the desired audio speed is half the original speed, then each frame is repeated, resulting in frames 202 $(FR_0...FR_m)$ of 1024 samples, where $FR_0 = FR_1 = F_0$, and $FR_2 = FR_3 = F_1$, etc. Additionally, m depends on the desired slow rate. In the example, where the desired audio speed is half the original speed, $m = 2n$. If, for example, the desired audio speed is two-thirds of the original speed, then every other frame is repeated, so frames 203 $(F_0...F_n)$ result in frames $(FR_0...FR_m)$, where $FR_0 = F_0$, $FR_1 = FR_2 = F_1$, $FR_3 = F_2$, $FR_4 = FR_5 = F_3$, etc., and $m = 3n/2$.

[0039] A window function WF is then applied to frames 202 $(FR_0...FR_m)$ to "smooth out" the samples and ensure that the resulting signal does not have any artifacts that may result from repeating each frame. The window function results in the windowed frames 204 $(WF_0... WF_L)$ of 1024 samples. The window function WF can one of many widely knows and used window functions, or can be designed to accommodate the design requirements of the system.

[0040] The windowed frames 204 $(WF_0... WF_L)$ of 1024 samples are then run through a digital-to-analog converter (DAC) to get an analog signal 201. The analog signal 201 is a longer version of the analog input signal 101 of **Fig. 4** (analog signal 201 and analog signal 101 are not equal). When the analog signal 201 is played at the same frequency as the original signal 101 of **Fig. 4**, the speed, in the

11

example with repeating each frame, is effectively half the speed at which the original audio was but the pitch remains the same, since the playback frequency remains unchanged. Hence, achieving a slower audio playback without affecting the pitch.

[0041]   **Figure 6** illustrates a flow diagram of an exemplary method for frequency-domain decoding of an audio signal, in accordance with an embodiment of the present invention. At a starting block 401, an input is received from the encoder directly, using a storage device, or through a communication medium. The input, which is coming from the encoder, is quantized and coded sets of frequency coefficients of a MDCT ($MDCT_0...MDCT_n$). At a next block 403 the input is inverse modified discrete cosine transformed, yielding sets ($wW_0...wW_n$) of 2048 windowed samples. An inverse window function is then applied to the windowed samples at a next block 405 producing the windows ($W_0...W_n$) each of which comprises 2048 samples. The windows are the result of overlapping frames ($F_0...F_n$), which may be obtained by inverse overlapping the windows ($W_0...W_n$) at a next block 407. Then depending on the rate at which the audio signal needs to be slowed down, the proper number of frames is replicated at a next block 409, as described above with reference to **Fig. 5**, resulting in the replicated frames ($FR_0...FR_m$).

[0042]   At a next block 410, a window function WF is applied to the frames ($FR_0...FR_m$) to "smooth out" the samples and ensure that the resulting signal does not have any artifacts that may result from repeating each frame. The window function results in the windowed frames ($WF_0... WF_L$). The window function WF can one of many widely knows

and used window functions, or can be designed to accommodate the design requirements of the system.

[0043]   The windowed frames (WF_0... WF_L) are then sent through the DAC at a next block 411 to produce the audio signal at the desired slower speed, with the same pitch as the original because the playback frequency is kept the same as the original signal.

[0044]   **Figure 7** illustrates a block diagram of an exemplary audio decoder, in accordance with an embodiment of the present invention.   The encoded audio signal is delivered from signal processor 301, and the advanced audio coding (AAC) bit-stream 303 is de-multiplexed by a bit-stream de-multiplexer 305.   This includes Huffman decoding 307, scale factor decoding 311, and decoding of side information used in tools such as mono/stereo 313, intensity stereo 317, TNS 319, and the filter bank 321.

[0045]   The sets of frequency coefficients 109 ($MDCT_0$...$MDCT_n$) of **Fig. 4** are decoded and copied to an output buffer in a sample fashion.   After Huffman decoding 307, an inverse quantizer 309 inverse quantizes each set of frequency coefficients 109 ($MDCT_0$...$MDCT_n$) by a 4/3-power nonlinearity.   The scale factors 311 are then used to scale sets of frequency coefficients 109 ($MDCT_0$...$MDCT_n$) by the quantizer step size.

[0046]   Additionally, tools including the mono/stereo 313, prediction 315, intensity stereo coupling 317, TNS 319, and filter bank 321 can apply further functions to the sets of frequency coefficients 109 ($MDCT_0$...$MDCT_n$).   The gain control 323 transforms the frequency coefficients 109 ($MDCT_0$...$MDCT_n$) into a time-domain audio signal.   The gain control 323 transforms the frequency coefficients 109 by

13

applying the IMDCT, the inverse window function, and inverse window overlap as explained above in reference to **Fig. 5**. If the signal is not compressed, then the IMDCT, the inverse window function, and the inverse window overlap are skipped, as shown in **Fig. 2**.

[0047] The output of the gain control 323, which is frames ($F_0...F_n$) such as, for example, frames 203 or frames 213, is then sent to the audio processing unit 325 for additional processing, playback, or storage. The audio processing unit 325 receives an input from a user regarding the speed at which the audio signal should be played or has access to a default value for the factor of slowing the audio signal at playback. The audio processing unit 325 then processes the audio signal according to the factor for slow playback by replicating the frames ($F_0...F_n$) at a rate consistent with the desired slow rate. For example, if the desired audio speed is half the original speed, then each frame is repeated, resulting in frames ($FR_0...FR_m$) such as, for example, frames 202 or frames 212, of 1024 samples, where $FR_0 = FR_1 = F_0$, and $FR_2 = FR_3 = F_1$, etc. The factor m depends on the desired slow rate. In the example, where the desired audio speed is half the original speed, $m = 2n$. If, for example, the desired audio speed is two-thirds of the original speed, then every other frame is repeated, so frames ($F_0...F_n$) result in frames ($FR_0...FR_m$), where $FR_0 = F_0$, $FR_1 = FR_2 = F_1$, $FR_3 = F_2$, $FR_4 = FR_5 = F_3$, etc., and $m = 3n/2$.

[0048] A window function WF is then applied to frames ($FR_0...FR_m$) to "smooth out" the samples and ensure that the resulting signal does not have any artifacts that may result from repeating each frame. The window function results in the windowed frames ($WF_0... WF_L$) such as, for example, frames 204 or frames 214, of 1024 samples. The window

14

function WF can one of many widely knows and used window functions, or can be designed to accommodate the design requirements of the system.

[0049]   At this point the signal is still in digital form, so the output of the audio processing unit 325 is run through a DAC 327, which converts the digital signal to an analog audio signal to be played through a speaker 329.

[0050]   In an embodiment of the present invention, the playback speed is pre-determined in the design of the decoder. In another embodiment of the present invention, the play back speed is entered by a user of the decoder, and varies accordingly.

[0051]   While the present invention has been described with reference to certain embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the scope of the present invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the present invention without departing from its scope. Therefore, it is intended that the present invention not be limited to the particular embodiment disclosed, but that the present invention will include all embodiments falling within the scope of the appended claims.